Vedic Post-Lexical Retroflexion: Opacity and Diachrony

Cathcart, Chundra

Vedic Post-Lexical Retroflexion: Opacity and Diachrony*

Chundra Aroor Cathcart
University of California, Berkeley

1 Introduction

Sanskrit possesses a well-known rule by which alveolar *n* becomes retroflex *ṇ* if *r*, *r̥* /ɻ̩/ or *ṣ* /ʂ/ precedes it and no coronal obstruent intervenes. In Vedic, Epic and Classical Sanskrit, this rule is completely regular within word boundaries. In Vedic, the oldest variety of Sanskrit, a process of *n*-retroflexion also occurs across word boundaries, albeit variably. This paper investigates the phonological factors conditioning Vedic post-lexical retroflexion of *n* (henceforth PLR), and the extent to which PLR conforms to theoretical claims regarding lexical phonology and morphology (LPM).

This paper demonstrates that Vedic PLR is an opaque, phonetically unnatural phenomenon. While it can be represented to some extent by models of LPM that allow for ordered levels of post-lexical phonology, this phenomenon presents problems for the assumption that lexeme-internal, structure-sensitive rules must precede post-lexical ones.

This paper advocates a diachronic explanation of PLR. I show that a model sensitive to diachronic factors better predicts the distribution of retroflex clitics than a model which assumes phonetic naturalness in prosodic phonology. At the same time, it is highly likely that the opaque and phonetically unnatural distribution of retroflex clitics led to the gradual under-generalization and loss of this already gradient phenomenon.

2 Vedic Retroflexion

In Vedic Sanskrit, a number of diachronic developments have yielded retroflex segments.

1. Proto-Indo-European **s* and its allophone **z* (found before voiced plosives) became Proto-Indo-Iranian **š* and **ž* when preceded by **r*, **r̥*, **u*, **k*, and **i* (the so-called RUKI rule). (PIIr **š* and **ž* were originally in complementary distribution, but underwent phonologization due to extension of morphemes containing **ž*.) In Pre-Vedic, **š* and **ž* underwent a context-free change to **ṣ* and **ẓ*. These segments triggered the retroflexion of following adjacent dental segments (as well as alveolar **n*). After triggering retroflexion, Pre-Vedic **ẓ* was deleted (with compensatory lengthening) before plosives. At some point during its development, PIIr **ž* was rhotacized before vowels and sonorant consonants. This development must have happened after Proto-Indic

*This paper has benefited greatly from the input of audiences at CLS 50 and SCIHS 3. Special thanks are due to Gary Holland for comments on an early draft. All errors are my own.

branched off from Proto-Indo-Iranian (against the idea that some rhotacistic developments took place in Proto-Indo-Iranian, see *inter alia* Kellens & Pirart 1990, 260), but it is not clear if it happened before the context-free retroflexion of **ž* to **ẓ* (what I term EARLY RHOTACISM), or afterward (LATE RHOTACISM, cf. Kobayashi 2004, 81).

2. Retroflexion of *n* is triggered by preceding *r*, *r̥* or *ṣ* so long as no coronal obstruent intervenes (the process is also blocked if *n* is directly followed by an adjacent dental). This development was synchronically quite productive and, as a lexeme-internal synchronic rule, is virtually exceptionless, and can be represented as follows:

$$(1) \quad n \rightarrow \underline{n} \,/\, \left\{ \begin{array}{c} r \\ \underset{\circ}{r} \\ \underline{s} \end{array} \right\} \neg \left[\begin{array}{c} +\text{cor} \\ +\text{obs} \end{array} \right]^{*} \_ \neg \left[\begin{array}{c} +\text{cor} \\ +\text{ant} \end{array} \right]$$

3. There are additional sporadic developments yielding retroflexes, probably due to dialectal influence, that are not relevant for the purposes of this paper.

As a lexeme-internal rule, Vedic (and generally, Sanskrit) *n*-retroflexion is widely discussed in the literature (Allen 1951; Allen 1953; Zwicky 1965; Collinge 1965). Some examples follow. A number of forms show this process morpheme-internally:

(2) a. *ráṇa-* 'battle' < **rana-*
b. *ū́rṇā-* 'wool' < **ūrnā-* < **Hu̯r̥Hnā-*
c. *kárṇa-* 'ear' < **karna-*

In addition, it is seen in morphophonemic alternations:

(3) a. /a-tr̥p-nu-vant-/ → *átr̥pṇuvant-* 'insatiable'
b. /rug-na-/ → *rugṇá-* 'broken'
c. /pari-māna-/ → *parimā́ṇa-* 'circumference'
d. /ramha-māna-/ → *raṃhamāṇa-* 'hastening'
e. /rek-nas-/ → *rékṇas-* 'wealth'
f. /parī-nas-/ → *párīṇas-* 'abundance'

In Vedic, an identical-looking *n*-retroflexion process operates variably across word boundaries as well:

(4) a. [$_{\omega}$ [$_{w}$ indra] [$_{w}$ nas]] 'Indra' (voc.) + 1pl pron.cltc. → *indra ṇo* ∼ *indra no*
b. [$_{\omega}$ [$_{w}$ aryamā] [$_{w}$ nas]] 'Aryaman' (voc.) + 1pl pron.cltc. → *aryamā́ ṇo* ∼ *aryamā́ nas*
c. [$_{\omega}$ [$_{w}$ pra] [$_{w}$ nas]] 'forth' (preverb) + 1pl pron.cltc. → *prá ṇo* ∼ *prá no*
d. [$_{\omega}$ [$_{w}$ uru] [$_{w}$ nas]] 'far' + 1pl pron.cltc. → *urú ṇaḥ* ∼ *urú naḥ*

PLR receives somewhat less attention in the literature than lexeme-internal retroflexion (though see Whitney 1879, 66, Wackernagel 1896 [1957], 191, Selkirk

1980). In the Ṛg and Atharva Vedas (the two oldest Vedic texts), a number of functional items like *nas/naḥ/no* (underlyingly /nas/, 1pl pron.cltc.), *na* (neg.ptcl.), *na* (simile ptcl.), *nu* (emph.ptcl) and *ena-* (proximal pron.) are affected, as well as lexical items such as *enas-* 'sin, transgression', though this is less common.

3 Phonetic Naturalness in Phonology

An ongoing debate revolves around whether phonetic naturalness or "plausibility" is an obligatory facet of phonology (Steriade 1997, Flemming 2001, etc.) or not (Bach & Harms 1972, Buckley 2000, Yu 2004, etc.). Much of this debate pertains to processes that take place within lexical boundaries; it is generally assumed that processes that occur across word boundaries consist of low-level ones, not far removed from phonetic implementation rules (or at least closer to them than lexical rules). Concerning the relationship between lexical and post-lexical phonology, Coetzee & Pater (2011:402) give the following schema, implicitly assumed by most generative phonologists:

(5) Lexicon
↓
Early (i.e., Lexical) Phonology
↓
Late (i.e., Post-lexical) Phonology
↓
Phonetic Implementation

Lexical phonology is prone to morphological and lexical exceptions. Post-lexical phonology operates at the phrasal level and is generally characterized as "exceptionless" due to its assumed independence from the language's morphosyntax and lexicon, but is often inhibited by extra-grammatical factors like speech rate, etc. The modular feed-forward architecture of phonology (Bermúdez-Otero 2014) envisions a "life cycle" of rules in which they start as phonetic stabilization at the phrasal level and subsequently make their way to higher domains of the grammar.

In the following sections, I investigate PLR from a quantitative standpoint, with an eye to the following questions: Is PLR the same process as lexeme-internal retroflexion, just at the phrasal as opposed to the word or stem level? To what extent is it predictable by the environments that condition lexeme-internal retroflexion?[1]

4 Corpus Study 1

I generated a corpus of all tokens of words in the Ṛg and Atharva (Śaunaka recension) Vedas that exhibit $n \sim ṇ$ allomorphy of the type described in §2 ($N = 6553$; $n = 6407$, $ṇ = 145$). For each token, a number of factors were incorporated into mixed-effects logistic regression models. Token type (clitic vs. particle, etc.), the word preceding the token, the word following the token, and the location of the token (by Book, by Veda) were assigned random intercepts.

[1] I assume that the operation of PLR is preceded by any syntactic movement that might occur.

Table 1: Counts: place of articulation by trigger (type)

	Trigger?		Adjacent?		Non-Adjacent?	
	N	Y	N	Y	N	Y
n	5350	1057	6056	351	5701	706
ṇ	0	145	119	26	26	119

I identified three predictors of interest. First, I wished to determine (1) whether or not the preceding word contained a synchronically VIABLE TRIGGER OF RETROFLEXION, i.e., whether *r*, *r̥* or *ṣ* preceded and no coronal obstruent intervened. Additionally, since distance between the trigger of a rule and affected segments can often determine whether or not a post-lexical process operates (particularly if intonational phrase boundaries intervene), I wished to determine whether triggers influencing PLR were (2) ADJACENT or (3) NON-ADJACENT. Non-adjacent triggers included both *ṣ* and *r*. Adjacent triggers were at the right word edge; the corpus only included adjacent triggers of the type *-r*, since word-final *-ṣ* is permitted only when it can syllabify with a voiceless plosive onset at the beginning of the next word. Three models were constructed with each factor of interest treated as a fixed effect; it was not possible to include all fixed effects in the same model, given their collinearity.

4.1 Results and Discussion

The significance of the fixed effects was determined using the likelihood ratio test. The presence of a viable trigger is a highly significant predictor ($\chi^2(1) = 90.2$, $p < .001$). The presence of a viable non-adjacent trigger is a highly significant predictor as well ($\chi^2(1) = 90.7$, $p < .001$). However, the presence of a viable adjacent trigger is only marginally significant ($\chi^2(1) = 3.89$, $p = .05$).

This last finding is odd for a number of reasons. First, post-lexical processes are common at word boundaries (e.g., nasal place assimilation in *gree*[m b]*ox*, *Sa*[m p]*ablo*, etc.). Furthermore, Vedic retroflexion can be triggered at any distance; hence, there seems to be no *a priori* reason that non-adjacent retroflexion should be better represented than adjacent retroflexion. If anything, we would expect the opposite.

Another strange pattern emerges from the data. In certain contexts where a viable trigger is present, PLR is categorically blocked. It consistently fails to operate across sequences which surface as *-o* and *-aṃ* (m.nom.sg. and m.acc.sg/neut.nom.-acc.sg. suffixes, respectively), as well as most instances of final *-r* (hence the marginally significant result for adjacent triggers of retroflexion seen above). For instance, we see sequences like RV 4.55.10c *índro no* '[may] Indra (nom.) [come] to us', RV 7.103.3 *pitáraṃ ná* 'as [son greets] father (acc.)', RV 5.11.4 *agnír no* '[may] Agni (nom.) [come to] our [sacrifice]', but never **índro ṇo*, **pitáraṃ ṇá*, **agnír ṇo*. It is easy enough to envision a post-lexical grammar where PLR is generally disfavored by final *-r* and blocked by final *-o* and *-ṃ*. But these constraints are otherwise not well motivated, and certainly not phonetically so — for instance, a round back vowel like [o] should perceptually enhance retroflexion, not block it (Hamann 2003:90–2).

Most of the contexts in which viably triggered retroflexion is either under-

or unrepresented involve the operation of so-called "external" sandhi rules at the word edge. These rules, which owe themselves to sound changes occurring sometime during the development of Pre-Vedic into Vedic, are conditioned by following voiced segments (both consonants and vowels):

(6) /-as/ → -o / _ [+voi][2]

(7) /-s/ → -r / $V_{[hi]}$ _ [+voi][3]

For underlying sequences like /indras nas/ and /agnis nas/, it is clear that external sandhi rules like exx. (6–7) are in a counter-feeding relationship with PLR. Forms like /agnis/ are underlyingly different from forms with etymological *-r* (e.g., *svàr-* 'sun', *vā́r-* 'waters', *āśír-* 'milk added to the Soma', *pitár-* 'father', etc.); forms of the latter type generally trigger PLR (cf. Hale 1990, 91). Since both external sandhi and PLR are technically speaking post-lexical rules, insensitive to structure, this opacity can be modeled synchronically in a theory that allows ordered levels of post-lexical phonology (Kaisse & Shaw 1985:24), as follows (clitics are not bracketed as independent words for any particular reason):

		[$_\omega$ [$_w$ indras] [$_w$ nas]]		[$_\omega$ [$_w$ agnis] [$_w$ nas]]		
(8)	1. PLR	indras	nas	agnis	nas	(blocked)
	2. sandhi	indro	no	agnir	no	

However, this rule ordering seems at odds with the idea that within the phonological component, more abstract phonological rules tend to feed lower-level phonetic processes (recall Coetzee & Pater, *loc. cit.*). External sandhi rules like those seen in (6–7) are telescoped, reflecting multiple layers of diachronic change. PLR, on the other hand, was most likely a low-level phonetic process, or at least not far removed from one.[4]

There is an additional ordering problem: across word boundaries, PLR and external sandhi are in a counter-feeding relationship. Word-internally, the same sandhi rules regularly feed *n*-retroflexion:

(9) a. /nis-nij-am/ 'robe' acc.sg. → nir-nij-am → *nirṇíjam*

b. /dus-nāmā/ 'having a bad name' nom.sg. → dur-nāmā → *durṇā́mā*

[2]A similar-looking development of PIIr **-as* to /-ah/ → *-ō* in Avestan, the most archaic Iranian language, is thought to have come about independently.

[3]It could be argued that underlying representations like /-is/ and /-us/ do not hew to the principle of Stampean Occultation, i.e., that URs should not violate surface-level phonotactics (Prince & Smolensky 1993:54), given that the sequences *is* and *us* are marginally attested in Vedic (due to the historical operation of the RUKI rule); these representations are used for the sake of parsimony.

[4]This idea preponderates in the phonological literature, given the (likely erroneous) assumption that Sanskrit *r* was "retroflex," given its class designation *mūrdhanya* 'cerebral'; a retroflex characterization of Sanskrit *r* is found only in the *Pāṇinīya Śikṣa*, a text apocryphally attributed to the eponymous Hindu grammarian, while older *Pratiśākhya* treatises "generally require an alveolar articulation" (Allen 1953:53–4). It is likely that some alveolar rhotics are highly compatible with perseverative retroflexion; cf. the results of (Narayanan *et al.* 1999:1999), who find counterclockwise (front-to-back) movement for the Tamil alveolar and postalveolar taps. This dynamic retraction could result in retroflexion of following segments so long as a coronal segment did not change the configuration of the tongue — whence a natural phonetic basis for PLR.

c. /rakṣas-hanam/ ‘killing *rākṣasa*s’ acc.sg. → rakṣo-hanam → *rakṣoháṇaṃ*[5]

This odd pattern is difficult to model in LPM. The /-as/ → *-o* and /-s/ → *-r* sandhi rules seen in (8–9), whether they operate within or across words, are essentially the same processes, differing only in terms of whether or not they feed an *n*-retroflexion rule. Additionally, none of the processes shown above displays any particular morphological sensitivity. Derivationally, we would have to posit two virtually identical sandhi and retroflexion rules that differ only in terms of whether they apply within or across lexical boundaries, like so:

			[$_\omega$ [$_w$ indras] [$_w$ nas]]	[$_\omega$ [$_w$ rakṣas-hanam]]
(10)	Lexical	sandhi 1	indras nas	rakṣo- hanam
		retroflexion 1	indras nas	rakṣo- haṇam
	Postlexical	retroflexion 2	indras nas	rakṣo- haṇam
		sandhi 2	indro no	rakṣo- haṇam

Positing two virtually identical sandhi rules is not only derivationally uneconomical: “internal” and “external” sandhi rules have long been acknowledged as the same phenomenon in the Indian grammatical tradition (Allen 1962:15).

If we were to treat sandhis 1 and 2 as the same process, we would have to posit two retroflexion rules, one structure-sensitive, applying only within lexeme boundaries. In order to achieve the transparent lexeme-internal pattern seen in the data, the structure-sensitive retroflexion rule would have to apply after the other retroflexion rule:

		[$_\omega$ [$_w$ indras] [$_w$ nas]]	[$_\omega$ [$_w$ rakṣas-hanam]]
(11)	retroflexion 1	indras nas	rakṣo- haṇam
	sandhi	indro no	rakṣo- haṇam
	retroflexion 2	indro no	rakṣo- haṇam

Another troubling fact is that the phonology seen in morphophonemic alternations (feeding) is different from that affecting clitics (counter-feeding), which violates the general view that a clitic is governed by the same phonology as its host (Zwicky 1985:286).

An Optimality Theoretic recasting of the rule-ordering scenarios shown above does not fare much better (space does not permit an exposition using tableaux). OT would more or less work in cases with underlying /-as/. We could envision a lexical grammar where correspondence constraints and markedness constraints outrank faithfulness constraints, thus achieving the transparent feeding relationship. In the post-lexical grammar, /o/ could (rather arbitrarily, as noted above) block the spread of retroflexion where it hasn’t already applied in the lexical grammar. However, a non-serial post-lexical OT grammar cannot model the opaque counter-feeding relationship between the rules /-s/ → *-r* / $V_{[hi]}$ _ [+voi] and PLR. As shown above, PLR is categorically blocked for sequences like *agnir no* ← /agnis nas/, but generally applies for forms with underlying /-r/, e.g., RV 1.148.1 *súvar ṇá* ‘like the sun’ ← /suvar na/.

[5]This compound must be a relatively late coinage, since the second member *han-* (< PIr **jʰan-*) historically began with a coronal obstruent that ought to block retroflexion.

Thus, we see a number of problems that Vedic PLR poses for traditional models of lexical and post-lexical phonology. Paradoxically, a highly transparent pattern is seen lexeme-internally, while an opaque one holds (to some extent) across lexeme boundaries. The post-lexical pattern resembles so-called "crazy rules" that are best explained as multiple layers of diachronic residue.

5 Corpus Study 2

Given these oddities, I generated a second corpus that differed from the first one in the following way: whereas previously all viable synchronic triggers were included in the model, I now took into account only viable triggers that would have operated at an earlier stage of Vedic when it would have more closely resembled Proto-Indo-Iranian. This involved "undoing" the following sound changes: *-o* < Proto-Indo-Iranian **-as*, *-hi* (2pl imperative marker) < PIIr **-dʰi*, *-ir* < PIIr **-iš* (vs. authentic *r*-stems like *svàr-* 'sun', *vā́r-* 'waters', *pitár-* 'father', etc.), etc.[6]

Essentially, a sequence like *índro no* would be treated as something like **indras nas*, and a sequence like *agnír no* would be treated as **agniš nas*. I admit here that a potential circularity arises from our representation of forms ending in underlying /-is/ or /-us/. The sandhi rule in (7) telescopes a diachronic development dating back to Proto-Indo-European, which may have taken one of two trajectories. Rhotacism could have taken place at a relatively late date, as follows (adapted from Kobayashi 2004, 152–3):

(12) PIE **-i/us* > PIIr **-i/uš* (RUKI rule) > Pre-Indic **-i/uṣ* > **-i/uẓ* (sandhi) > Vedic *-i/ur* (LATE RHOTACISM)

Alternatively, rhotacism postdated the split of Pre-Indic from the larger body of Proto-Indo-Iranian, but took place before retroflexion:

(13) PIE **-i/us* > PIIr **-i/uš* (RUKI rule) > **-i/už* (sandhi) > **-i/ur* (EARLY RHOTACISM) > Vedic *-i/ur*

Regardless of whether rhotacism took place early or was directly preceded at a diachronic stage by **ṣ* or **ẓ*, at some stage in time, these forms contained sounds that could be interpreted as historical triggers (either **ẓ* or **r*). It may seem question-begging to exclude these forms based on their behavior (while continuing to include non-adjacent *ṣ*), but the fact remains that they are underlyingly and historically distinct from forms with etymological *-r*. The somewhat paradoxical behavior of non-etymological *-r* is discussed below, along with possible explanations for it.[7]

[6] Additional changes concerning sounds that would have blocked PLR, e.g., *h* < PIIr **ɟʰ*, **ʄʰ*, etc., were not relevant for the purposes of this corpus study.

[7] An additional problem has not yet been fully resolved: it is well known that Vedic poets made various attempts to avoid hiatus (Gunkel & Ryan 2011). In some places where a case form of a noun would result in a sequence of identical vowels, it is replaced with a "quirky" case form that resolves the hiatus. A striking example is AVŚ 12.5.72 *agnír enaṃ*, where *agnír* is morphologically nominative, but syntactically vocative **agne*. I have not been able to take into account all such instances, and furthermore, it is not clear when during the composition of the Vedas this sort of change took place (if even all at once).

Table 2: Counts: place of articulation by historical trigger (type)

	Hist trigger?		Adjacent?		Non-Adjacent?	
	N	Y	N	Y	N	Y
n	5824	583	6392	15	5839	568
ṇ	1	144	120	25	26	119

The same random intercepts were assigned as in the first corpus study. The presence of a HISTORICALLY VIABLE TRIGGER and whether it was ADJACENT or NON-ADJACENT were modeled as fixed effects in three separate mixed-effects logistic regression models.

5.1 Results and Discussion

The significance of the fixed effects was again determined by carrying out likelihood-ratio tests for each model. The presence of a historically viable trigger is a highly significant predictor ($\chi^2(1) = 157.9$, $p < .001$). The presence of a historically viable non-adjacent trigger is also a highly significant predictor ($\chi^2(1) = 116$, $p < .001$). The presence of a historically viable adjacent trigger is now highly significant as well ($\chi^2(1) = 195$, $p < .001$), given that these are only found in forms with etymological final *-r* (as opposed to those with final *-r* via external sandhi), and are no longer grossly overestimated. There is a lone instance of a non-historical adjacent *-r* triggering retroflexion in the following word (AVŚ 12.2.12c *mucyámāno nír éṇasó 'mog asmā́m̐ áśastyāḥ* 'released from the transgression, [Agni] freed us from the curse'), an outlier in which *nír* (underlying /nis/, an indeclinable particle meaning 'out, away') has spread retroflexion to following /enas-/. I have no particularly exciting explanation for this single aberrant datum — it simply looks like /nis/ and /enas-/, two separate words, were analyzed as one heteromorphemic word, allowing a transparent relationship between sandhi and retroflexion.[8]

6 Implications for Phonological Theory

I have provided data above which shows that the lexeme-internal and lexeme-external manifestations of rhotic retroflexion seen in Vedic are at considerable odds with models of lexical phonology and morphology which conceive of lexeme-internal processes as prone to blocking and lexical exceptions (due to historical residue, difficulties in learnability, etc.), but processes at the phrasal level as exceptionless (e.g., Kaisse & Shaw 1985, Kiparsky 1985, etc.). Additionally, this data is incompatible with feed-forward models of phonologization which view phonetic implementation rules as remaining active at the phrasal level as they move upward to the levels of the word and stem (Bermúdez-Otero 2014). Here, the process of retroflexion was blocked by sandhi at the phrasal level, but continued to be productive (and fed by sandhi) lexeme-internally. The lexeme-external pattern is transparent, while the post-lexical one is highly opaque. For the relationship between PLR

[8]It is worth noting the doctrinal complexities of this verse, part of a hymn about the fire-god Agni's consumption of corpses, in which he is variously (and paradoxically) exalted and deprecated. This may have caused some confusion surrounding the use of the word *enas-*.

Table 3: Forms triggering PLR

		n	*ṇ*
árṣā	pour (2sg imper.)	0	1
aryamā́	Aryaman (voc.)	3	1
āśī́r	milk added to the Soma	0	1
bráhmā	prayer (neut.acc.pl.)	0	1
ind(a)ra	Indra (voc.)	5	7
kṣárā	flow (2sg imper)	0	1
nḗṣi	lead (2sg imper.)	0	1
pári	around (preverb)	1	20
párṣi	rescue (2sg imper.)	0	1
prá	forth (preverb)	21	39
purupriyā́	dear to many (m.sg.inst.)	0	1
rákṣā	protect (2sg imper.)	0	5
śakrá	epithet of Indra (voc.)	0	1
sávaneṣu	pressing (loc.pl.)	0	2
śíkṣā	guide (2sg imper.)	0	4
ṣmā	emph.ptcl.	0	1
ṣŭ	good (pref.)	1	21
sutéṣu	Soma (loc.pl.)	0	2
súvar	sun (nom.sg.)	0	20
urú	far	2	5
uruṣyā́	protect (2sg imper.)	0	5
vā́r	waters (neut.nom.sg.)	0	4
nír /nis/	away, out of	1	1

and sandhi, the opacity is so great that PLR is in fact confined to 23 inflected forms of various semantic fields (though a noticeable number pertain to the pressing of Soma), listed in Table 3[9] — well on its way to a lexically dependent phenomenon.

It is of course essential to consider whether this truly is a natural phenomenon worthy of challenging the architecture concerning processes at different modules of the grammar. Vedic, Epic and Classical Sanskrit have been treated by many scholars as natural language, but studies also call for the need to distinguish between natural language and poetic grammar (Gunkel & Ryan 2011). The diachrony of the Vedas is not quite the same as the diachrony of the vernacular that existed alongside them; after all, the Vedas were passed down orally between generations for millennia. The domain in which sound change is usually thought to occur is within words, not hymns or sentences. At the same time, language is capable of evolving in units that are larger than single words, as in the grammaticalization of collocations. Thus, it is perfectly feasible to envision salient host-clitic pairs evolving jointly over time and maintaining (if only for a while) an opaque pattern, though this remains, to my knowledge, to be seen borne out by data from other (non-poetic) languages.

[9] A legitimate concern regarding this phenomenon is whether or not PLR is confined to collocations of high frequency or some sort of ritual salience. The use of random intercepts by each type of preceding word was intended to control for this possibility, though no information was recorded about each preceding word's ritual salience. The fact that PLR operated is none the less relevant.

7 Implications for Vedic Studies

The findings of this study have the potential to inform a number of debates about Vedic phonetics and phonology. An open question is whether Vedic contained one rhotic, or two. Catford (2001:181–3) provides some evidence for the claim that non-syllabic *r* was alveolar in most environments, while syllabic *r̥* was "molar." He argues also that final *-r* caused by sandhi was a fricative or approximant (and that the sandhi ending *-o* came about due to confusion between **-au* and **-ar* [aɪ]? < **-az* < **-as*, though we are not forced to accept this last idea). The results of this study show that *-r* due to sandhi virtually never induces PLR (except in the lone case of AVŚ 12.2.12c given above). It is well known that Swedish and Norwegian dialects with uvular [ʀ] do not undergo rhotic retroflexion unless they have borrowed a retroflexion rule from a neighboring dialect with alveolar [ɾ] (Svantesson 2000; Stausland Johnsen 2012). Thus, *r* caused by rhotacism of PIE **-s* may have been some sort of dorsal rhotic contrasting with an alveolar etymological *r* at the time when PLR was productive. This scenario makes this most sense if rhotacism is a relatively early development, bypassing the context-free retroflexion of PIIr **š/ž* (cf. ex. 13).

An unresolved question concerns final *-ṃ*, which categorically blocks PLR of initial *n-* in the following word (e.g., RV 7.103.3 *pitáraṃ ná*, but never **pitáraṃ ṇá*). This sound, called *anusvāra* (usually glossed as something like 'subordinate sound' or 'following sound'), alternates with word-final *-m*, appearing pre-nasally as well as in other contexts. Phonetic treatises provide an ambiguous picture of the *anusvāra*'s phonetic realization (Allen 1951:39–47). It is thought to be homorganic with following plosives, given the non-occurrence of heterorganic sequences like *[mk], *[mt], etc. in Vedic. And while heterorganic nasal sequences are permitted in Vedic (*mn* is quite common, *mṇ* less so), *ṃ*$_w$] [$_w$*n* is generally thought to have been realized as [nː] at the surface level (Allen 1962:80).

If *ṃ*$_w$] [$_w$*n* is realized as geminate [nː], the picture is still vague as to whether it should be affected by PLR, because it is not entirely clear whether *-n(-)n-* is subject to retroflexion lexeme-internally. This seems not to be the case in the earliest language (when we look at sequences derived from underlying /-d-n-/), e.g., [$_\omega$ [$_w$ yajus] [$_w$ skad-na-m]] → RV *yajuḥ ṣkannám* 'the [first] *Yaju*s that had descended', /ad$^{\text{h}}$i-skad-na-/ → TS *ádhiṣkanna-* 'covered', vs. /aː-ṭṛd-na/ → TS *ātr̥ṇṇa-* 'pierced', /ni-sad-na-/ → ŚB, TS *niṣaṇṇa-* 'seated' (Whitney 1879:67). If we analyze these patterns using ordered rules, it appears that in the earlier forms, a rule /d-n/ → *-nn-* and PLR are in a counter-feeding relationship, but in the later forms, they are in a feeding relationship. While it is possible that the failure of PLR across word-final *-ṃ* has something to do with the fact that *ṃ*$_w$] [$_w$*n* is realized as geminate [nː], there is no straightforward phonological reason as to why it would be blocked (we could appeal to geminate inalterability preventing *-nn-* from undergoing retroflexion, but the later forms with *-ṇṇ-* shown above fly in the face of this idea), much less one with *a priori* grounding in observable, well-established sandhi rules (as in the case of forms with underlying /d-n/, since *d* blocks the retroflexion rule).

8 Conclusion

Diachronic explanation is often invoked in the case of arbitrary or strange-looking synchronic alternations, usually within the domain of lexeme-internal phonology. This paper has provided evidence of opaque, phonetically unnatural prosodic phonology in the archaic poetic language of the Vedas, challenging the generally accepted idea that "late" phonology is necessarily phonetically natural. Instead, it appears that fossilization of historical processes happens in the prosodic word as well as the lexical word; just as structural gaps and exceptions to rules can arise lexeme-internally, similar patterns in certain domains of the prosodic word can be due to residue from an earlier diachronic stage. It seems that learners were able to contend with a transparent pattern of lexeme-internal morphophonology and an opaque rule operating across word boundaries.

At the same time, this opacity does seem to have created some difficulty for learners. Table 2 shows that roughly only a fourth of "historical" triggers induced PLR in the following form. It is likely that this non-operation of PLR is not solely due to the sorts of extragrammatical factors one would expect to impede prosodic phonology, but also the undergeneralization of the unnatural pattern governing post-lexical *n*-retroflexion. It appears that this undergeneralization was sufficiently advanced by the time of the earliest Vedic attestations, though PLR continues to operate variably through the Vedic and post-Vedic periods. PLR is virtually absent by the era of Epic and Classical Sanskrit (though the latter variety of Sanskrit is subject to strict prescriptive grammatical rules). Tentatively, the results of this paper support a view of phonology where at the level of competence, learners have the capacity to acquire patterns largely independent of phonetics, but usage-level constraints can potentially lead to the abandonment of highly opaque alternations.

References

Allen, W. S. 1951. Some prosodic aspects of retroflexion and aspiration in Sanskrit. *BSOAS* 13.939–946.

—— 1953. *Phonetics in Ancient India*, volume 1 of *London Oriental Series*. London: Oxford University Press.

—— 1962. *Sandhi*. Number 17 in Janua Linguarum. The Hague: Mouton.

Bach, Emmon, & Robert Harms. 1972. How do languages get crazy rules? In *Linguistic Change and Generative Theory*, ed. by R. P. Stockwell & R. K. S. Macaulay, 1–21. Bloomington, IN: Indiana University Press.

Bermúdez-Otero, Ricardo. 2014. Amphichronic explanation and the life cycle of phonological processes. Preprint. In *The Oxford handbook of historical phonology*, ed. by Patrick Honeybone & Joseph Salmons. Oxford: Oxford University Press.

Buckley, Eugene. 2000. On the naturalness of unnatural rules. In *Proc. 2nd Workshop on American Indigenous Languages*, volume 9 of *UCSB Working Papers in Linguistics*.

Catford, John C. 2001. On Rs, rhotacism and paleophony. *JIPA* 31.171–185.

Coetzee, Andries W., & Joe Pater. 2011. The place of variation in phonological theory. In *The Handbook of Phonological Theory*, ed. by John Goldsmith, Jason Riggle, & Alan C. L. Yu, 401–434. Malden, MA: Blackwell, 2 edition.

Collinge, Neville E. 1965. Some linguistic paradoxes. *J. Ling* 1.1–12.
Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.7–44.
Gunkel, Dieter, & Kevin Ryan. 2011. Hiatus avoidance and metrification in the Rigveda. In *Proceedings of the 22nd Annual UCLA Indo-European Conference*, 53–68. Bremen: Dr. Ute Hempen Verlag.
Hale, Mark. 1990. Preliminaries to the study of the relationship between sandhi and syntax in the language of the Rigveda. *Münchener Studien zur Sprachwissenschaft* 51.77–96.
Hamann, Silke. *The phonetics and phonology of retroflexion*. University of Utrecht dissertation.
Kaisse, Ellen M., & Patricia A. Shaw. 1985. On the theory of lexical phonology. *Phonology Yearbook* 2.1–30.
Kellens, Jean, & Eric Pirart. 1990. *Les textes vieil-avestiques. Vol II: Répertoires grammaticaux et lexique*. Wiesbaden: Dr. Ludwig Reichert Verlag.
Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology Yearbook* 2.85–138.
Kobayashi, Masato. 2004. *Historical Phonology of Old Indo-Aryan Consonants*. Tokyo: Fujiwara.
Lubotsky, Alexander. 1997. *A Ṛgvedic word concordance*, volume 82 of *American Oriental Series*. New Haven: American Oriental Society.
Narayanan, Shrikanth, Dani Byrd, & Abigail Kaun. 1999. Geometry, kinematics, and acoustics of Tamil liquid consonants. *J. Acoust. Soc. Am.* 106.1993–2007.
Prince, Alan, & Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report RuCCS-TR-2, Rutgers University Center for Cognitive Science, New Brunswick, NJ.
Selkirk, Elisabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In *Juncture: a collection of original papers*, ed. by Mark Aronoff & Mary-Louise Kean, volume 7 of *Studia Linguistica et Philologica*, 107–130. Saratoga, CA: Anma Libri.
Stausland Johnsen, Sverre. 2012. A diachronic account of phonological unnaturalness. *Phonology* 29.505–531.
Steriade, Donca, 1997. Phonetics in phonology: The case of laryngeal neutralization. University of California, Los Angeles.
Svantesson, Jan-Olof. 2000. Phonology of a southern Swedish idiolect. *Lund University Working Papers* 49.156–159.
Wackernagel, Jakob. 1896 [1957]. *Altindische Grammatik. Vol. 1: Lautlehre*. Göttingen: Vandenhoeck & Ruprecht.
Whitney, William Dwight. 1879. *A Sanskrit grammar*. Leipzig: Breitkopf & Härtel.
Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: phonetics and history. *Language* 80.73–97.
Zwicky, Arnold M. *Topics in Sanskrit Phonology*. MIT dissertation.
—— 1985. Clitics and particles. *Language* 61.283–305.

Chundra Aroor Cathcart
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall

Berkeley, CA 94720–2650
chundra@berkeley.edu